

OBSERVED DIFFERENCES IN RHYTHM BETWEEN PERFORMANCES OF CLASSICAL AND JAZZ VIOLIN STUDENTS

Enric Guaus, Oriol Saña

Escola Superior de Música de Catalunya
{enric.guaus, oriol.sana}@esmuc.cat

Quim Llimona

Universitat Pompeu Fabra
quim.llimona01@estudiant.upf.edu

ABSTRACT

The aim of this paper is to present a case study that highlights some differences between violin students from the classical and jazz traditions. This work is part of a broader interdisciplinary research that studies whether classical violin students with jazz music background have more control on the tempo in their performances. Because of the artistic nature of music, it is difficult to establish a unique criteria about what this control on the tempo means. The case study here presented quantifies this by analyzing which student performances are closer to some given references (i.e. professional violinists). We focus on the rhythmic relationships of multimodal data recorded in different sessions by different students, analyzed using traditional statistical and MIR techniques. In this paper, we show the criteria for collecting data, the low level descriptors computed for different streams, and the statistical techniques used to determine the performance comparisons. Finally, we provide some tendencies showing that, for this case study, the differences between performances from students from different traditions really exist.

1. INTRODUCTION

In the last centuries, learning musical disciplines has been based on the personal relationship between the teacher and the student. Pedagogues have been collecting and organizing such a long experience, specially from the classical music tradition, for proposing learning curricula in conservatories and music schools. Nevertheless, because of the artistic nature of music, it is really difficult to establish an objective measure between performances from different students, so, it is very difficult to objectively analyze the pros and cons of different proposed programs.

In general, a musician is able to adapt the performance of a given score in order to achieve certain musical and emotional effects, that is, provide an *expressive musical performance*. There exists a huge literature for the analysis of expressive musical performances. Widmer [1] provides a good overview on this topic. Under our point of view, one of the most relevant contributions is the Performance Worm for the analysis of performances by Dixon [2]. It

shows the evolution of tempo and perceived loudness information in a 2D space in real time, with a decreasing brightness according to a negative exponential function to show past information. Saunders [3] analyzed the playing styles from different pianists using (beat-level) tempo and (beat-level) loudness information. In the opposite direction, different systems have been developed to allow machines create more expressive music, which are summarized by Kirke [4]. Then, according to the literature, most of the studies related to expressive performance are based on loudness and rhythmic properties of music.

This research is part of a PhD thesis on art history and musicology. Its aim is to present evidences in differences of performances for violin students from jazz and classical traditions, in terms of rhythm. We decided focusing on rhythm of music because is one of the key aspects to work with classical violin students, and it is coherent with the existing literature. For that, we propose a methodology based on multimodal data collection from different pieces, students and sessions and analyze it using state-of-the-art techniques from statistics and Music Information Retrieval (MIR) fields.

This paper is organized as follows: Section 2 explains the experimental setup for data acquisition. Section 3 shows the statistical analysis we used for further discussion in Section 4. Finally, the conclusions and future work are presented in Section 5.

2. EXPERIMENTAL SETUP

The aim of this setup is to capture rhythmic properties of the proposed performances. It is specially designed to make our future analysis independent of the played violin, the played piece, the particular student and the particular playing conditions of a specific session. We are only interested on the musical tradition of the two groups of students: those coming from the jazz tradition and those coming from the classical tradition.

2.1 Participants

We had the collaboration of 8 violin students (Students A...H) from the Escola Superior de Msica de Catalunya (ESMUC), in Barcelona. Some of them are enrolled in classical music courses (subjects A, G) while others are enrolled both in classical and jazz music courses (subjects B, C, D, E, F, H). We also recorded two well known professional violinists as a reference, one from the classical

tradition (subject *I*) and the other from the jazz tradition (subject *J*).

2.2 Exercises

We asked students to perform different pieces from the classical and jazz tradition as in a concert situation. Pieces were selected according to their rhythmic complexity, according to the criteria of both classical and jazz tradition professional violinist.

W. A. Mozart . Symphony n.38 in Eb Maj, 1st. movement, KV 543: Rhythmic patterns with sixteenth notes and some eighth notes in between. This excerpt presents high rhythmic regularity.

R. Strauss . Don Juan, op. 20, excerpt: Rhythmic excerpts that are developed through out the piece. There exists small variations on the melody but rhythm remains almost constant.

R. Schumann . Symphony n. 2 in C Maj, Scherzo, excerpt: Rhythmic complexity is higher than the two previous pieces. This excerpt does not present a specific rhythmic pattern.

Schreiber . Rhythm exercise proposed by jazz violin professor Andreas Schreiber, from Anton Bruckner University, Linz.

Charlier . Rhythm exercise proposed by drums professor Andr Charlier, from Le centre des musiques Didier Lockwood, Dammarie-Ls-Lys, France.

Gustorff . Rhythm exercise proposed by jazz violin teacher Michael Gustorff from ArtEZ Conservatory Arnhem, The Netherlands.

All students played classical tradition pieces but only jazz students were able to perform jazz tradition pieces. Because of that, for the further analysis, we only use classical tradition exercises and we only compute distances from student performances to the professional violinist from the classical tradition.

2.3 Sessions

We follow the students through 10 sessions in one trimester, from September to December 2011, in which they had to play all the exercises. With that, we want to make results independent of particular playing conditions in a specific session. Reference violinists were asked to play as in a concert situation, and they were recorded only once.

2.4 Data acquisition

For all the exercises, students and sessions, we created a multimodal collection with video, audio and bow-body relative position information. Position sensors were mounted on a unique violin. We asked students to perform twice, first with their own violin to obtain maximum richness in expressivity recording audio and video streams, and a second performance on the violin and bow with all the sensors

attached. In this last case, all the participants performed on the same violin. We also recorded audio and video streams using both violins. In this research, we only include position and audio streams.

2.4.1 Audio

We recorded audio stream for the two types of violin for each exercise, student and session. We collected audio from (a) ambient microphone located at 2m far away from the violin, and clip-on microphone to capture timbre properties of the violin, and (b) a pickup attached to the bridge to obtain more precise and room independent data from the violin. We only use pickup information in our analysis.

2.4.2 Position

As detailed in previous research, the acquisition of gesture related data can be done using position sensors attached to the violin [5]. Specifically, we use the Polhemus¹ system, a six degrees of freedom electromagnetic tracker providing information on localization and orientation of a sensor with respect to a source. We use two sensors, one attached to the bow and the other attached to the violin obtaining a complete representation of their relative movement. From all the available data, we focus on the following streams that can be directly computed: Bow position, bow force and bow velocity.

This data is sampled at $sr = 240Hz$ and converted to audio at $sr = 22050Hz$ to allow feature extraction, as will be described in the following section. Video, audio and position streams are partly available under a Creative Commons License [6].

3. ANALYSIS

Right now, we collected the audio and position streams for each exercise, student, session and violin type. Now, we compute a set of rhythmic and amplitude descriptors from the collected streams and search for the dependence between them and the groups of students.

3.1 Feature extraction

We start computing descriptors from the audio recorded from the pickup (1 stream @ $sr = 22050Hz$) and from the position data from the sensors attached to the violin (3 streams @ $sr = 240Hz$). Data from Polhemus sensors is resampled to $sr = 22050Hz$. After some preliminary experiments, descriptors obtained through this resampling were determined to be related with rhythm, even assuming what we compute is not exactly the expected descriptor. We compute two sets of descriptors using MIR toolbox for Matlab [7]: (a) a set of compact descriptors for each audio excerpt including length, beatedness, event density, tempo estimation (using both autocorrelation and spectral implementations), pulse clarity, and low energy; and (b) a *bag of frames* set of descriptors including onsets, attack time and attack slope².

¹ <http://www.polhemus.com/>

² Attack time and attack slope are considered timbric descriptors, but we also include them in our analysis.

Descriptor	Student		Session		Exercise		Type	
length	9.20e-03	xxx	2.43e-01	-	1.69e-49	xxx	6.39e-01	-
beatedness	3.79e-01	-	1.52e-01	-	1.45e-15	xxx	2.01e-01	-
event density	3.54e-03	xx	1.49e-02	x	9.78e-27	xxx	6.42e-01	-
tempo estimation (autoc)	1.20e-01	-	5.16e-01	-	5.93e-18	-	6.68e-01	-
tempo estimation (spec)	9.14e-02	-	9.21e-01	-	7.98e-36	xxx	7.21e-01	-
pulse clarity	1.31e-02	x	4.47e-01	-	4.63e-99	xxx	5.24e-01	-
low energy	2.81e-02	x	6.93e-01	-	5.25e-89	xxx	4.96e-01	-
onsets	1.96e-01	-	4.25e-01	-	2.04e-01	-	1.44e-10	xxx
attack time	2.80e-03	xx	7.81e-01	-	2.24e-01	-	3.84e-01	-
attack slope	9.92e-05	xxx	2.30e-01	-	7.30e-01	-	7.17e-02	-

Table 1. Results of 1-way ANOVA analysis of the differences between the students and the classic tradition reference with the computed descriptors from the audio from the pickup.

Descriptor	Student		Session		Exercise	
length	2.67e-01	-	7.73e-01	-	7.84e-34	xxx
beatedness	8.86e-01	-	8.52e-01	-	1.15e-02	x
event density	9.84e-01	-	9.08e-01	-	1.41e-66	xxx
tempo estimation (autoc)	5.35e-01	-	8.52e-01	-	6.72e-23	xxx
tempo estimation (spec)	8.33e-01	-	8.66e-01	-	2.71e-13	xxx
pulse clarity	6.24e-01	-	6.35e-01	-	8.35e-09	xxx
low energy	7.59e-01	-	9.26e-01	-	2.15e-76	xxx
onsets	7.41e-01	-	9.52e-01	-	3.19e-10	xxx
attack time	1.14e-01	-	3.45e-01	-	9.05e-02	-
attack slope	6.70e-01	-	9.50e-01	-	2.87e-02	x

Table 2. Results of 1-way ANOVA analysis of the differences between the students and the classic tradition reference with the computed descriptors from the bow displacement.

Descriptor	Student		Session		Exercise	
length	2.67e-01	-	7.73e-01	-	7.84e-34	-
beatedness	1.74e-01	-	7.08e-02	-	3.51e-02	x
event density	3.39e-01	-	8.13e-01	-	3.27e-51	xxx
tempo estimation (autoc)	3.46e-01	-	9.51e-01	-	1.07e-13	xxx
tempo estimation (spec)	7.36e-01	-	7.10e-01	-	4.24e-13	xxx
pulse clarity	3.99e-01	-	8.24e-01	-	2.45e-25	xxx
low energy	5.93e-01	-	4.52e-01	-	4.70e-26	xxx
onsets	7.21e-01	-	8.72e-01	-	2.53e-11	xxx
attack time	8.47e-01	-	9.75e-01	-	3.20e-15	xxx
attack slope	9.76e-01	-	7.59e-01	-	2.14e-18	xxx

Table 3. Results of 1-way ANOVA analysis of the differences between the students and the classic tradition reference with the computed descriptors from the bow force.

Descriptor	Student		Session		Exercise	
length	2.67e-01	-	7.73e-01	-	7.84e-34	xxx
beatedness	1.85e-01	-	5.84e-01	-	1.65e-02	x
event density	7.53e-01	-	8.95e-01	-	6.27e-40	xxx
tempo estimation (autoc)	2.38e-01	-	9.75e-01	-	1.08e-08	xxx
tempo estimation (spec)	4.57e-01	-	2.92e-01	-	1.74e-17	xxx
pulse clarity	6.65e-01	-	4.23e-01	-	1.82e-14	xxx
low energy	6.84e-01	-	9.38e-01	-	1.07e-51	xxx
onsets	6.56e-01	-	2.84e-01	-	2.85e-04	xxx
attack time	9.52e-01	-	1.17e-01	-	2.14e-01	-
attack slope	7.52e-01	-	1.68e-01	-	4.08e-01	-

Table 4. Results of 1-way ANOVA analysis of the differences between the students and the classic tradition reference with the computed descriptors from the the bow velocity.

Descriptor	Pickup		Bow disp.		Bow force		Bow vel.	
length	9.08e-05	xxx	8.70e-01	-	8.70e-01	-	8.70e-01	-
beatedness	6.82e-03	xx	9.62e-02	-	5.66e-01	-	3.27e-04	xxx
event density	5.03e-01	-	4.27e-01	-	2.14e-02	x	2.34e-01	-
tempo estimation (autoc)	6.30e-04	xxx	9.39e-04	xxx	5.64e-04	xxx	3.39e-01	-
tempo estimation (spec)	2.75e-03	xx	9.71e-04	xxx	3.40e-01	-	5.49e-01	-
pulse clarity	5.91e-10	xxx	2.66e-01	-	8.44e-05	xxx	8.08e-04	xxx
low energy	5.04e-17	xxx	3.52e-01	-	1.11e-01	-	2.98e-02	x
onsets	1.90e-01	x	6.07e-01	-	1.17e-01	-	1.22e-01	-
attack time	1.76e-02	x	3.67e-02	x	3.53e-01	-	4.61e-01	-
attack slope	4.33e-02	x	3.94e-01	-	1.37e-01	-	4.92e-01	-

Table 5. Results of 2-way ANOVA analysis (student and exercise) of the differences between the students and the classic tradition reference with the computed descriptors from different streams

As mentioned in Section 1, according to pedagogic criteria, our work is based on the existing differences between the student performances (participants $A \dots H$) and the professional references (participants I, J). As detailed above, after the analysis of the recorded data, we observed that all the students played the exercises from the classical tradition with a high quality, while only those with jazz background played properly the exercises from the jazz tradition. Then, all the comparisons are computed in relation to the classical tradition professional violinist (participant I).

For the first set of (compact) descriptors, we compute the euclidean distance between the obtained descriptors of all the recordings from the students and their relative value from the professional performance.

For the frame-based descriptors, as the student and reference streams are not aligned, we use Dynamic time warping (DTW) [8] which also proved to be robust in gesture data [9]. Specifically, we use the total cost of warping path as a distance measure between two streams.

In summary, we have a set of descriptors related to the rhythmic distance between students and the reference for 4 streams of data (one from audio and three from position).

3.2 Statistical analysis

One-way Analysis of variance (ANOVA) is used to test the null-hypothesis within each variable, assuming that sampled population is normally distributed. Null hypothesis are defined as follows:

$$H_0: \text{Descriptor } X \text{ do not influence the definition of variable } Y.$$

being X one of the rhythmic descriptors detailed in Section 3.1, and Y one of the four variables in our study (student, session, exercise, and type). Results shown in Tables 1, 2, 3, 4 represent the probability of null hypothesis being true. Then, we consider that descriptor X is representative for $p(H_0) \leq 0.05$. We also include a graphic marker to detect when the descriptor has a certain influence according to the following criteria: (a) – for $0.01 \leq p(H_0) \leq 0.05$, no influence; (b) x for $0.001 \leq p(H_0) < 0.01$, small influence; (c) xx for $0.0001 \leq p(H_0) < 0.001$, medium influence; (d) xxx for $p(H_0) < 0.0001$, strong influence.

It is also interesting to analyze results of two-way ANOVA analysis for the student and exercise variables of our study. Results are shown in Table 5, also including graphical markers.

4. DISCUSSION

As detailed in the Section 3.2, Table 1, 2, 3 and 4 show the results of the 1-way ANOVA analysis of the differences between the performances played by the students and the reference for different streams and descriptors. *Type* variable is only taken into account in the analysis of pickup data because Polhemus streams are only recorded using one violin, as described in Section 2.4.2. Nevertheless, as the null hypothesis can not be rejected for most of the descriptors, we conclude that the violin type has no influence in our analysis. Moreover, the probabilities of null hypotheses for *Session* variable are also high. The null hypotheses can not be rejected, then, we conclude that the *Session* variable has no influence in our analysis.

Focusing on the *Exercise* and *Student* variables in Tables 1, 2, 3, and 4, we observe a high dependence of the *Exercise* variable in most of the descriptors and streams, as expected. Our goal is to analyze the behavior of the students. Table 5 shows the results of the two-way ANOVA analysis for *Student* and *Exercise* joint variables (Note how, in this table, columns represent different streams, not variables, for space restrictions). Null hypotheses can be rejected for different descriptors and variables, but we observe a high accumulation of 'xxx' graphic markers for tempo estimation (auto-correlation) and pulse clarity descriptors³. We guess that these descriptors are the best to explain differences between the two groups of students.

Moreover, according to Tables 1...5, we observe how the most representative stream is the audio recorded from the pickup. For that, from now to the end, we focus only on this stream.

Assuming ANOVA shows these descriptors present some statistically significant dependency with the two groups of students, we can go back to the original data and analyze

³ Pulse clarity is considered as a high-level musical dimension that conveys how easily in a given musical piece, or a particular moment during that piece, listeners can perceive the underlying rhythmic or metrical pulsation [10].

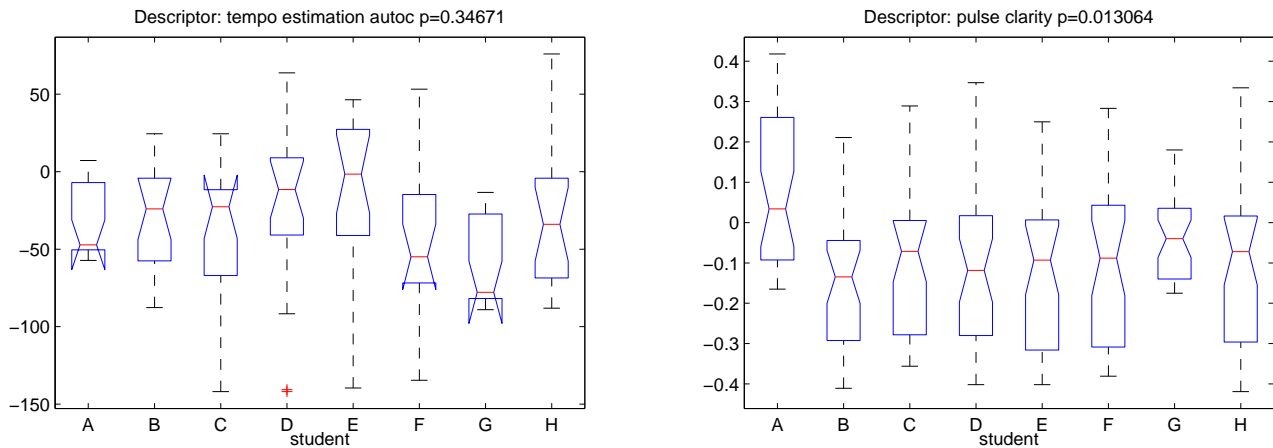


Figure 1. 1-way anova analysis plots for (a) tempo estimation (auto-correlation) descriptor on student variable, using bow-force estimation stream, and (b) pulse clarity descriptor on student variable, using pickup stream.

its behavior. Figure 1 shows the statistics for tempo estimation (auto-correlation) and pulse clarity descriptors (those who presented a high dependence in the ANOVA analysis) with respect to the classical tradition reference. Even with the *Exercise* variable information scrambled in these plots, we observe how student *A* and *G* present a different behavior with respect to the other ones. As described in Section 2.1, students *A* and *G* are those without jazz musical background.

Focusing on the tempo estimation (auto-correlation) shown in Figure 1 (a), we can derive some partial conclusions:

- Mean of the relative tempo estimation for students from the jazz tradition are far from the professional violinist, except for the participant *F*. Assuming a negative value of the difference means that the student plays faster than the reference, we observe a tendency on classical students playing faster than the reference.
- The lower limit (25th. percentile) of the relative tempo estimation for students from classical tradition are close to their mean. This could mean classical tradition students are more stable in their tempo.

Focusing on the pulse clarity shown in Figure 1 (b), we can derive some partial conclusions:

- Mean values of the relative pulse clarity for students from the classical tradition are closer to zero. We deduce the pulse-clarity for students from the classical tradition is closer to the professional violinist.
- Mean values of the relative pulse clarity for students from the jazz tradition are far and negative. Assuming a negative value of the difference means that the student plays with a higher pulse clarity than the reference, we could deduce that students from the jazz tradition show a clearer pulse than the reference.
- The lower limit (25th. percentile) of the samples for students with jazz background is lower than the

lower limit of the samples for students with classical background. As in the previous case, assuming a negative value of the difference means that the student plays with a higher pulse clarity than the reference, we could deduce that students from the jazz tradition show a clearer pulse than the reference.

It is not the goal of this paper to pedagogically define what does it mean *to perform better*, but we guess that, in our scenario, students with jazz musical background can be objectively identified in terms of tempo and pulse clarity with respect to those students without this background. For all, we conclude that the two groups of students can be objectively identified.

5. CONCLUSION

In this paper, we presented a case study for the comparison of musical performances in terms of rhythm of two groups of students. Specifically we proposed a methodology to determine which parameters may best identify rhythmic properties of performances carried out by a given set of students under specific conditions, based on multimodal data, and analyzed whether they are closer to a given reference. The novelty of this methodology is the obtention of rhythmic properties related to a group of students instead of a specific student, piece, session, or violin. Data from the pickup resulted being more effective than gesture data from the position sensors. Pulse clarity and tempo estimation showed to be the descriptors that have a major influence in the student behavior. Then, by analyzing them in detail, we observe how the two separable groups they provide coincide with the groups of students defined by their musical background, as shown in Figure 1. This can be a controversial conclusion for pedagogic and artistic research. In order to make these conclusions more general, our next step is to increase the number of subjects to analyze, including more scores, participants and instruments.

Acknowledgments

The research leading these results has received funding from the European Union Seventh Framework Programme FP7 / 2007-2013 through the PHENICX project under grant agreement n 601166.

6. REFERENCES

- [1] G. Widmer and W. Goebel, "Computational models of expressive music performance: The state of the art," *Journal of New Music Research*, vol. 33, no. 3, pp. 203–216, 2004.
- [2] S. Dixon, W. Goebel, and G. Widmer, "The performance worm: Real time visualization of expression based on langner's tempo-loudness animation," in *Proceedings of the International Computer Music Conference (ICMC)*, Gteborg, Sweden, 2002, pp. 361–364.
- [3] C. Saunders, D. Hardoon, J. Shawe-taylor, and W. Gerhard, "Using string kernels to identify famous performers from their playing style," in *Proceedings of the 15th European Conference on Machine Learning (ECML)*, 2004.
- [4] A. Kirke and E. Reck Miranda, "A survey of computer systems for expressive music performance?" *ACM Surveys*, vol. 42, no. 1, 2009.
- [5] E. Maestre, M. Blaauw, J. Bonada, E. Guaus, and A. Perez, "Statistical modeling of bowing control applied to violin sound synthesis," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 4, pp. 855–871, May 2010.
- [6] O. Mayor, J. Llop, and E. Maestre, "Repovizz: A multimodal on-line database and browsing tool for music performance research," in *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR)*, Miami, USA, 2011.
- [7] O. Lartillot and P. Toiviainen, "A matlab toolbox for musical feature extraction from audio," in *Proceedings of the International Conference on Digital Audio Effects*, Bordeaux, France, 2007.
- [8] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 26, no. 1, pp. 43–49, 1978.
- [9] M. Muller, "Efficient content-based retrieval of motion capture data," *ACM Transactions on Graphics*, vol. 24, no. 3, 2005.
- [10] O. Lartillot, T. Eerola, P. Toiviainen, and J. Fornari, "Multi-feature modeling of pulse clarity: Design, validation and optimization," in *Proceedings of the 9th International Society for Music Information Retrieval Conference (ISMIR)*, Philadelphia, PA, USA, 2008, pp. 521–526.